Understanding the Regularity and Privacy Issues in Human Mobility Research

Song Gao
Geospatial Data Science Lab
University of Wisconsin-Madison

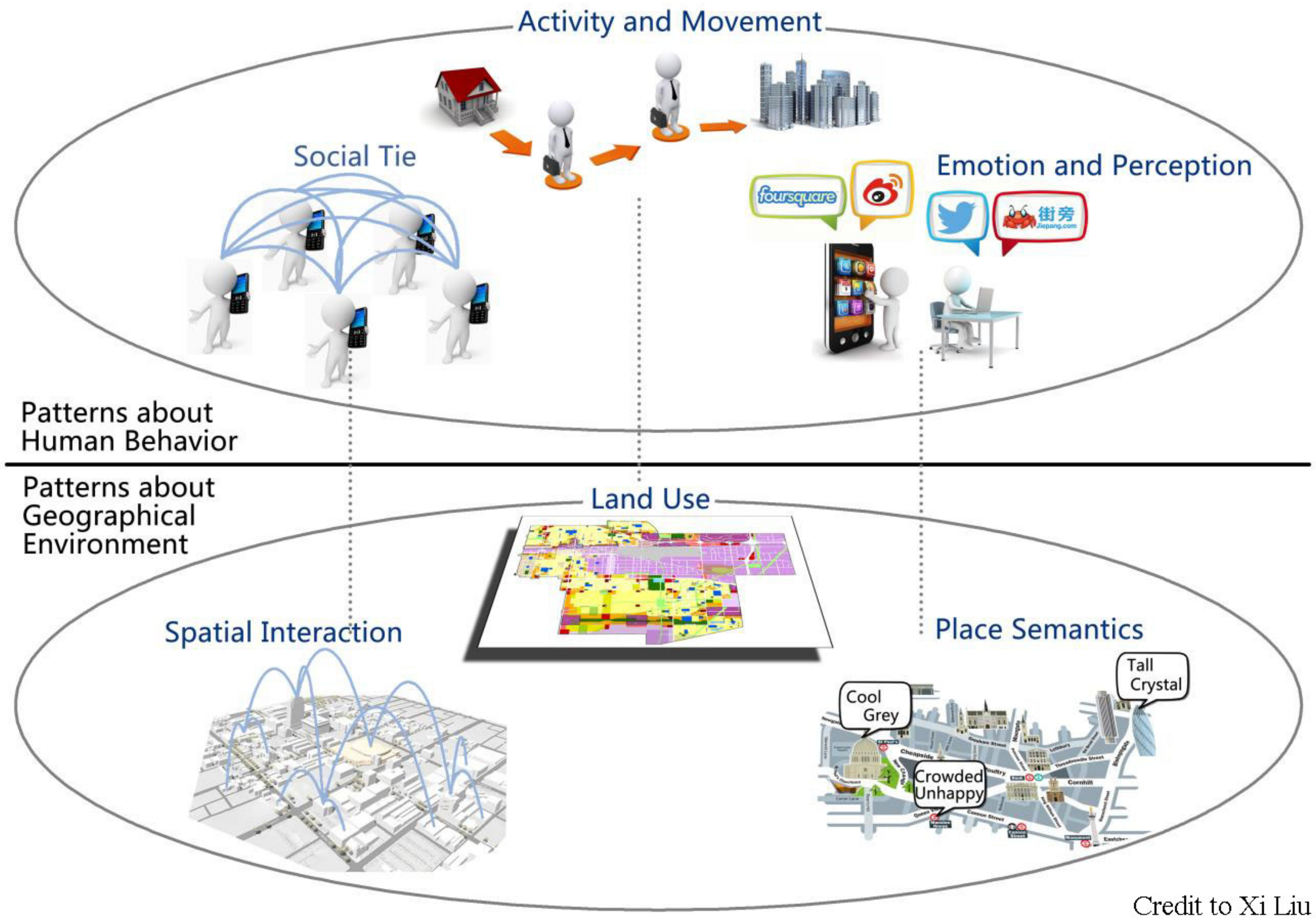
Email: song.gao@wisc.edu

https://geods.geography.wisc.edu/

Human Mobility Analytics: Data Sources and Regularity

Data-Driven Human Mobility Research

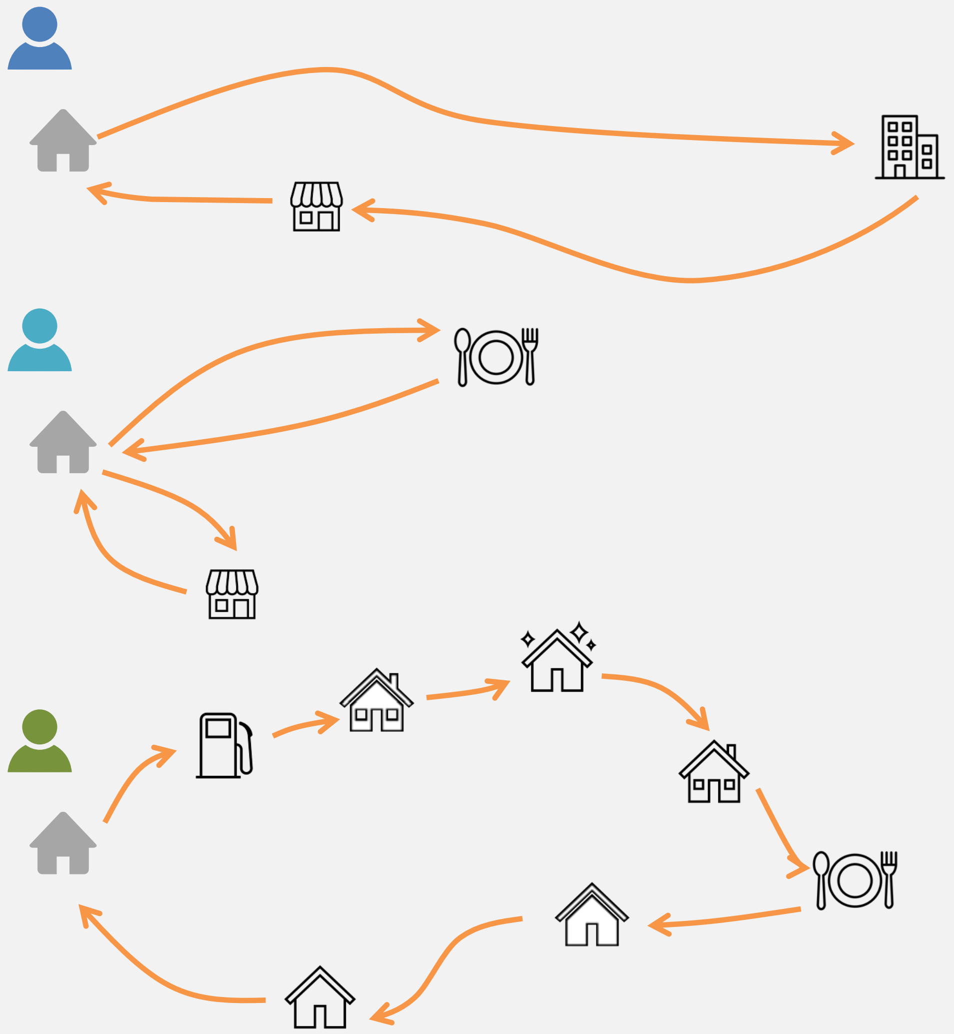
- Trip Surveys
- GPS trajectories
- Call detail records (CDRs)
- Social media
- Location-based apps
- Telematics data



Credit to Xi Liu

Regularity in daily human mobility patterns

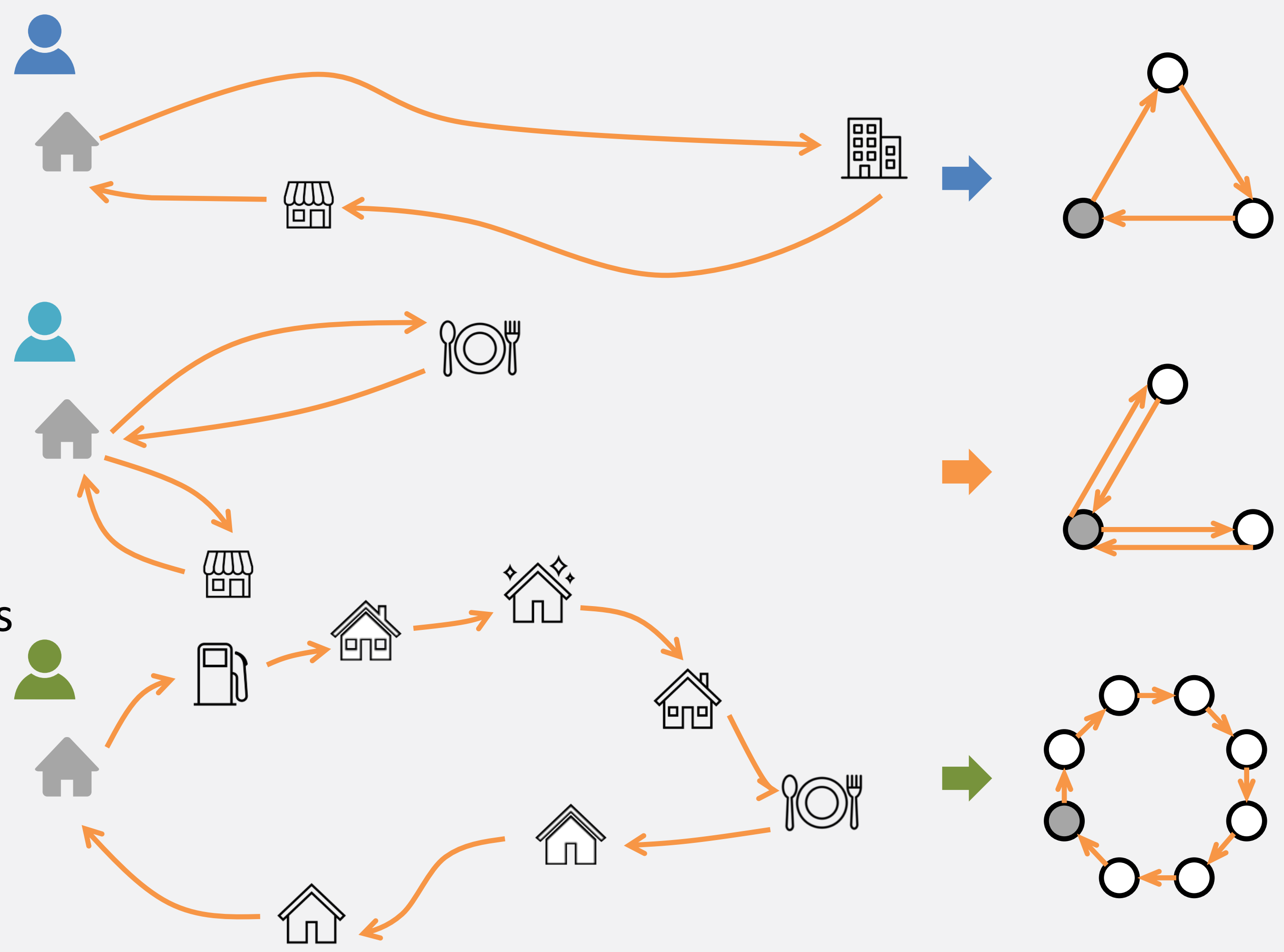
- Daily Routines



- **Underlying mechanism**
  - Needs/Constraints
    - Work
    - Rest
    - Social
  - Flexibility
    - Occupation
    - Exploration
    - Preferences

Mobility Motif: Location-based Motif

- Network Representation for Daily Routines
  - Node
    - Visited locations
  - Directed Edge
    - Trip between locations

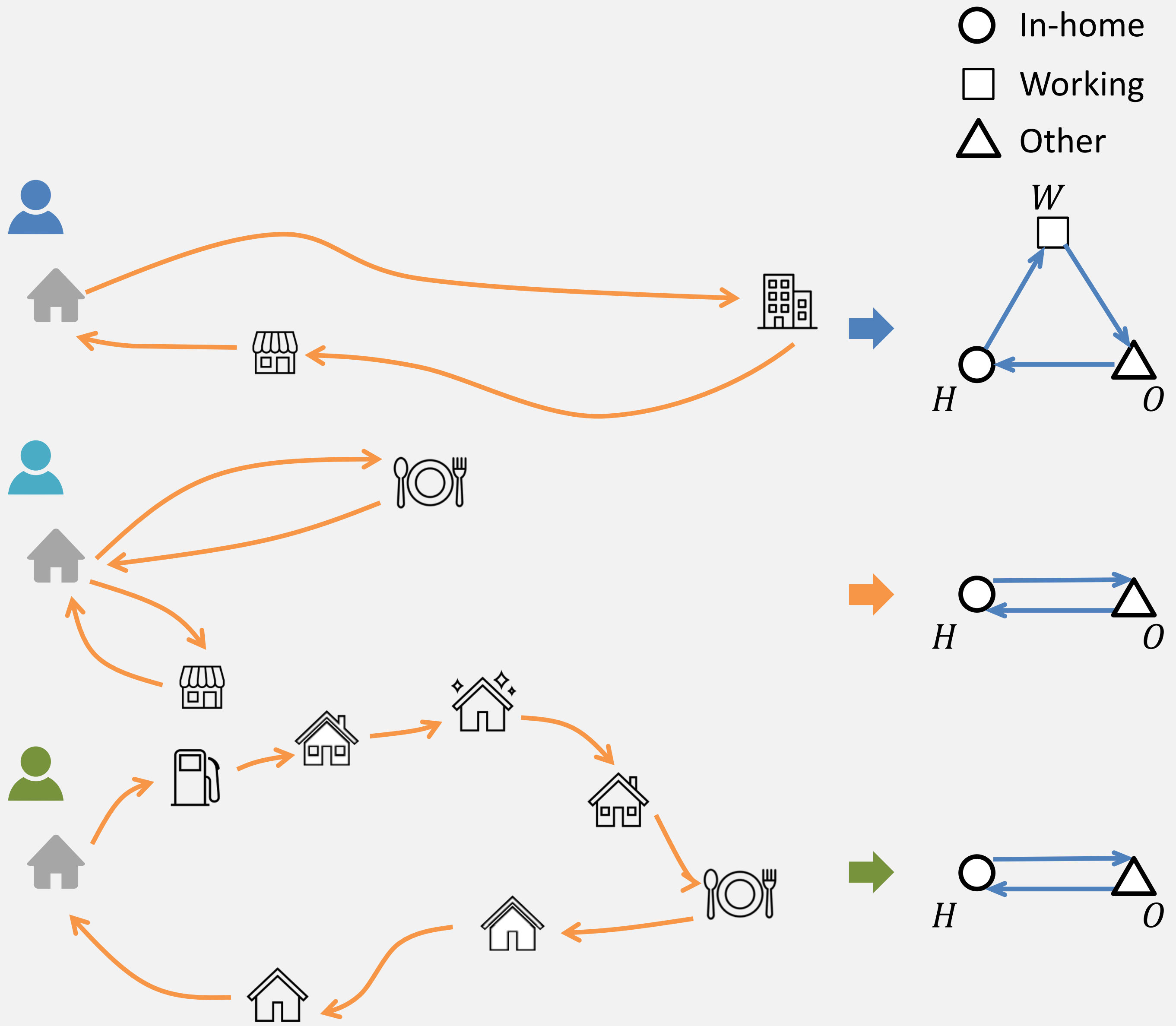


- Motif
  - Discard node attributes
    - Name, dwelling time, number of visits
  - Discard edge attributes
    - Number of trips, distance, travel time
  - Focus on **the network structure**

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. Journal of The Royal Society Interface, 10(84), 20130246.

Mobility Motif: Activity-based Motif

- Motivation
  - Explore how **activity types** affect location-based motif choices
- Modification
  - Node
    - Activity type (i.e., in-home, work, and others)
  - Directed Edge
    - Transition between activities

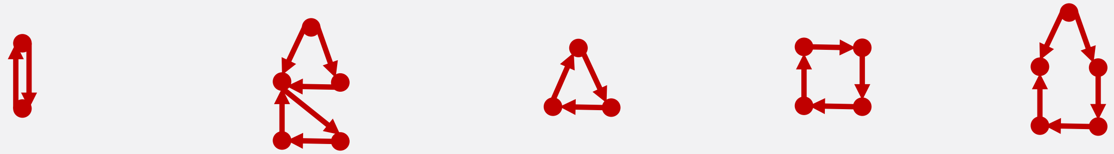


Cao, J., Li, Q., Tu, W., & Wang, F. (2019). Characterizing preferred motif choices and distance impacts. *Plos one*, *14*(4), e0215242.

Regularity in daily human mobility patterns

- Typical Users
  - User 1: 19 weekdays within one month

Location-based motif



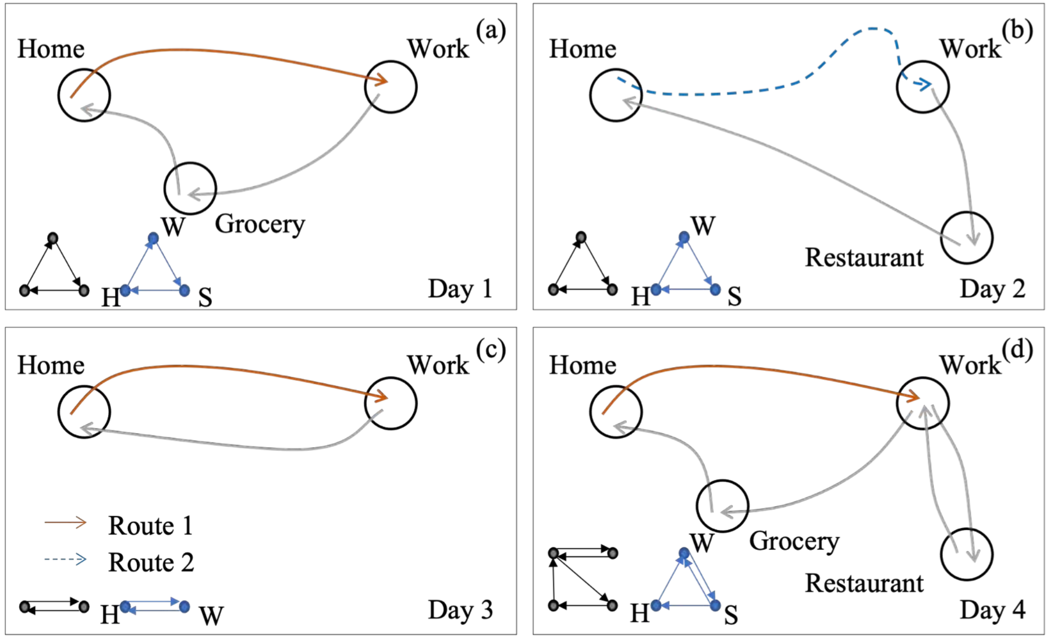
3 days | 1 day | 9 days | 3 days | 3 days

Activity-based motif

4 days | 15 day

Regularity in daily human mobility patterns

- Regularity in Location Motif
- Regularity in Activity Motif
- Regularity in Routing



Feature		Description
Location-related	N_{Loc}	The number of unique stay locations
	F_{Loc}	The frequency of the most frequently visited location
	$P_{Loc}(Home)$	The percentage of the total dwelling time at home
	$P_{Loc}(Work)$	The percentage of the total dwelling time at work place
	N_{Trip}	The total number of trips
	N_{OD}	The number of unique OD pairs
	F_{OD}	The frequency of the most frequently traveled OD pair
	$P_{Trip}(Home)$	The percentage of home-related trips
	$P_{Trip}(Comm)$	The percentage of commuting trips
	S_{Loc}	The actual entropy introduced in (Song et al., 2010)
Motif-related	N_{LM}	The number of unique LMs
	F_{LM}	The frequency of the most frequently used LM
	S_{LM}	The entropy of LMs
	N_{AM}	The number of unique AMs
	F_{AM}	The frequency of the most frequently used AM
	S_{AM}	The entropy of AMs
Route-related	F_R	The frequency of the most frequently used route
	S_R	The entropy of all routes, weighted by OD frequency

Ji, Gao, et al. Rethinking the regularity in mobility patterns of personal vehicle drivers: A multi-city comparison using a feature engineering approach. (2023), *Transactions in GIS*.

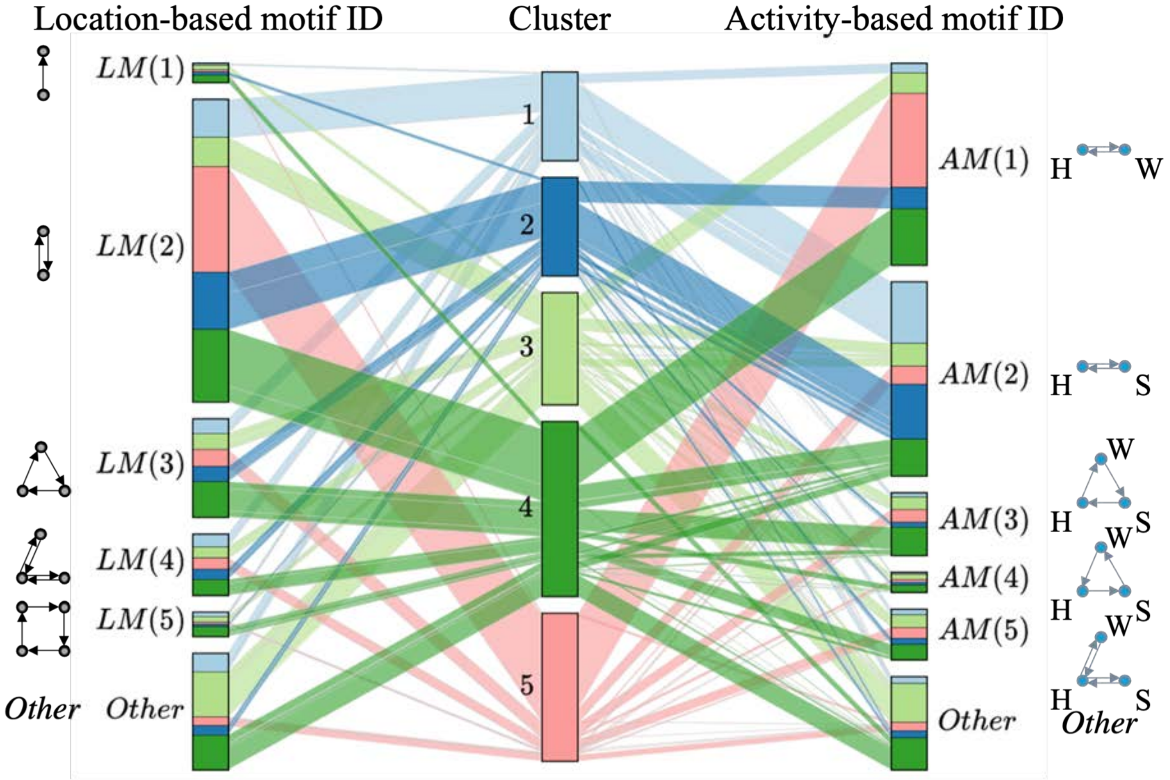
User Profiling/Clustering

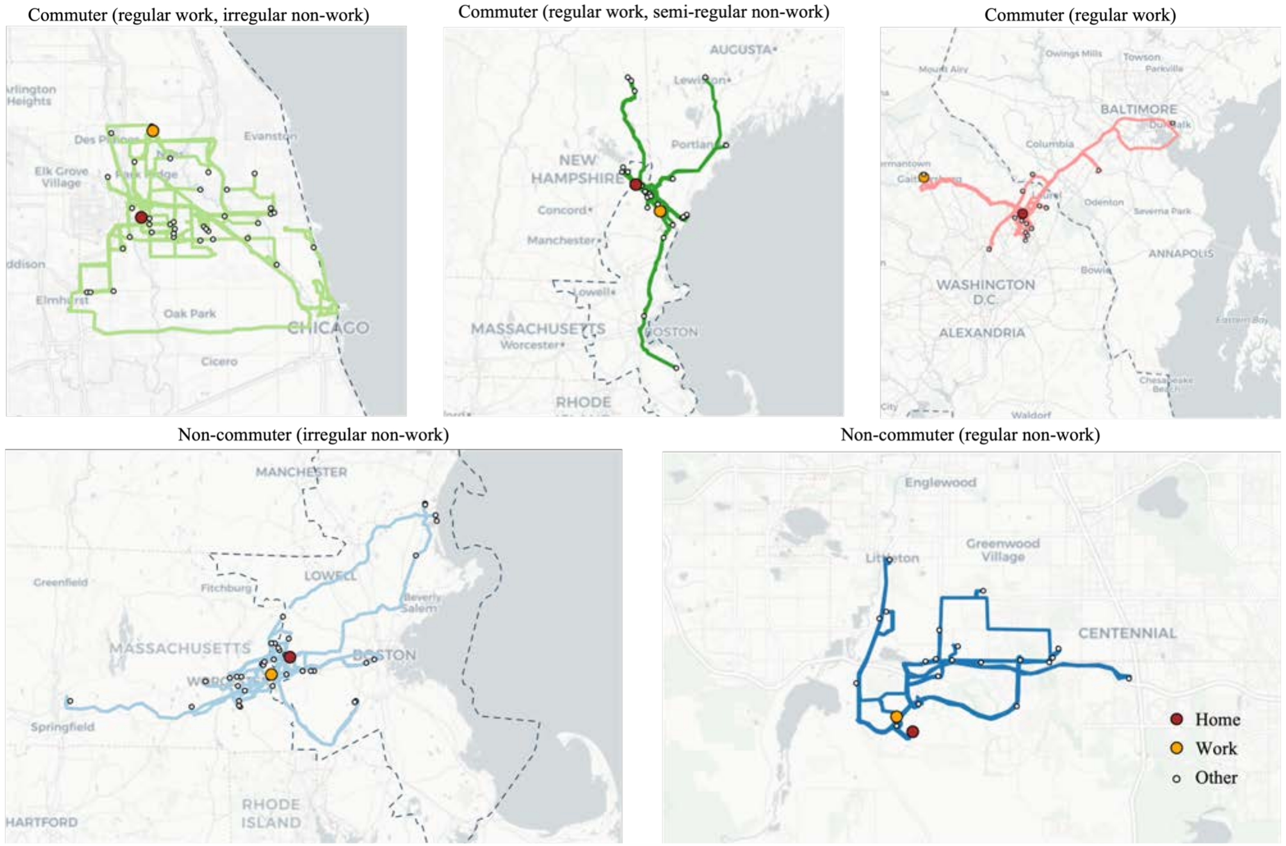
Transactions in GIS

Rethinking the regularity in mobility patterns of personal vehicle drivers: A multi-city comparison using a feature engineering approach

Yuhan Ji[1] | Song Gao[1,*] | Tam Huynh[2] | Chris Scheele[2] | James Triveri[2] | Jacob Kruse[1] | Collin Bennett[2] | Yicheng Wen[2]

Cluster 1: Gig drivers
Cluster 2: Homebodies
Cluster 3: Movers
Cluster 4: Typical drivers
Cluster 5: Work-focused commuters



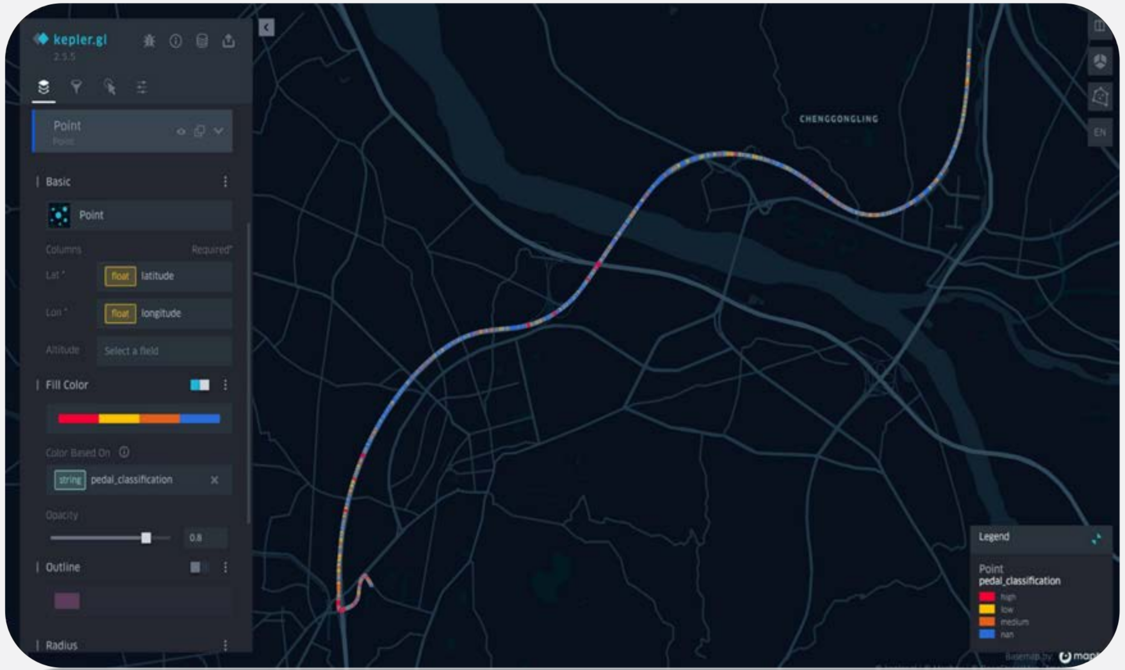


Ji, Gao, et al. (2023), Transactions in GIS.

GeoAI for Location Privacy Protection

Privacy Concerns

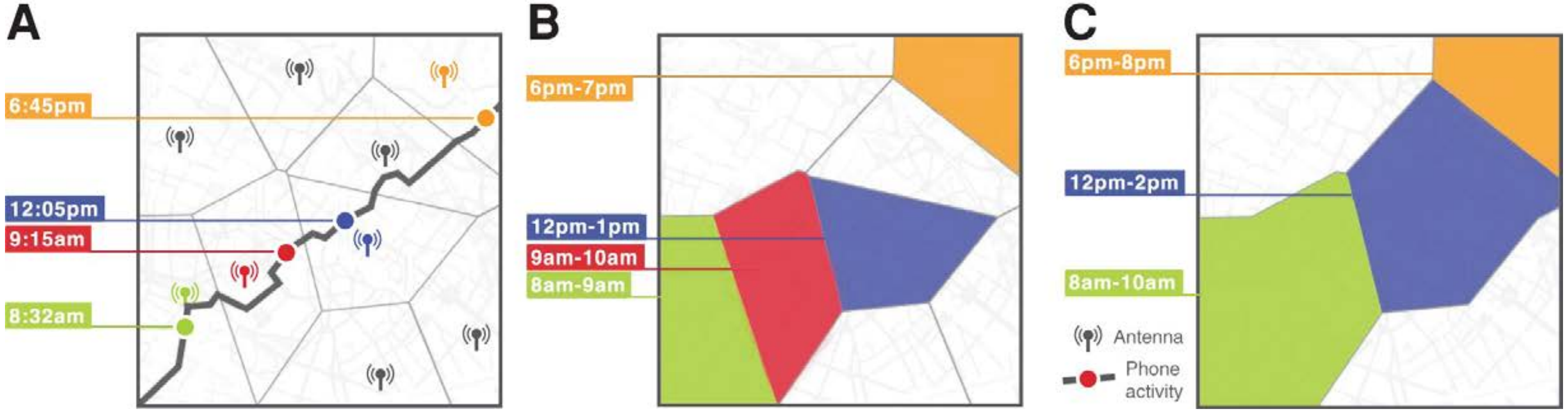
- **Geoprivacy** refers to an individual's rights to prevent the disclosure of personal sensitive locations including but not limit to their home, workplace, daily activities, or travel trips.



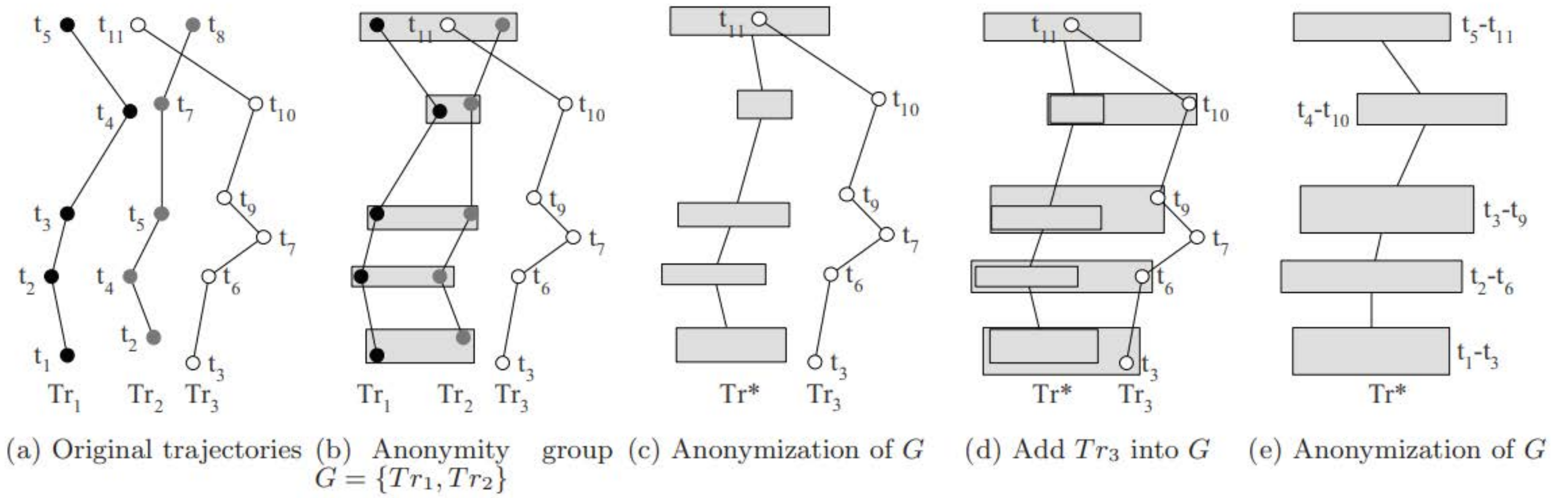
Gao, S., Rao, J., Liu, X., Kang, Y., Huang, Q., & App, J. (2019). Exploring the effectiveness of geomasking techniques for protecting the geoprivacy of Twitter users. *Journal of Spatial Information Science*, 2019(19), 105-129.

Geoprivacy Protection Methods

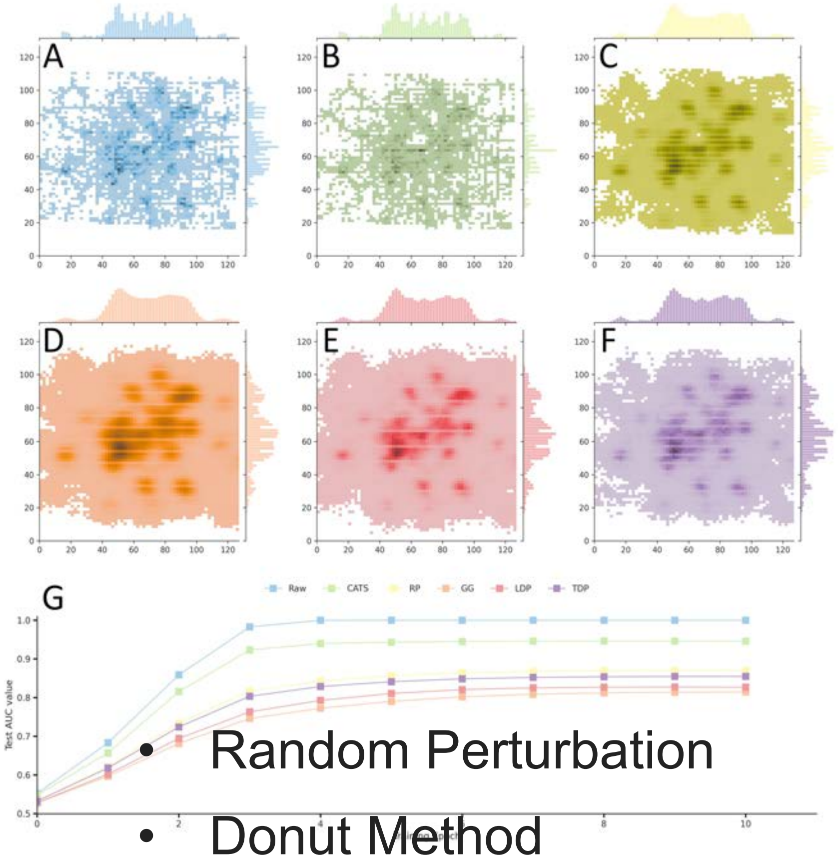
Aggregation / Geomasking / Grouping / Differential Privacy



Aggregate into geographic or administrative units (Nature, 2009)



Grouping and mixing trajectories to achieve k-anonymity (SIGSPATIAL, 2012)

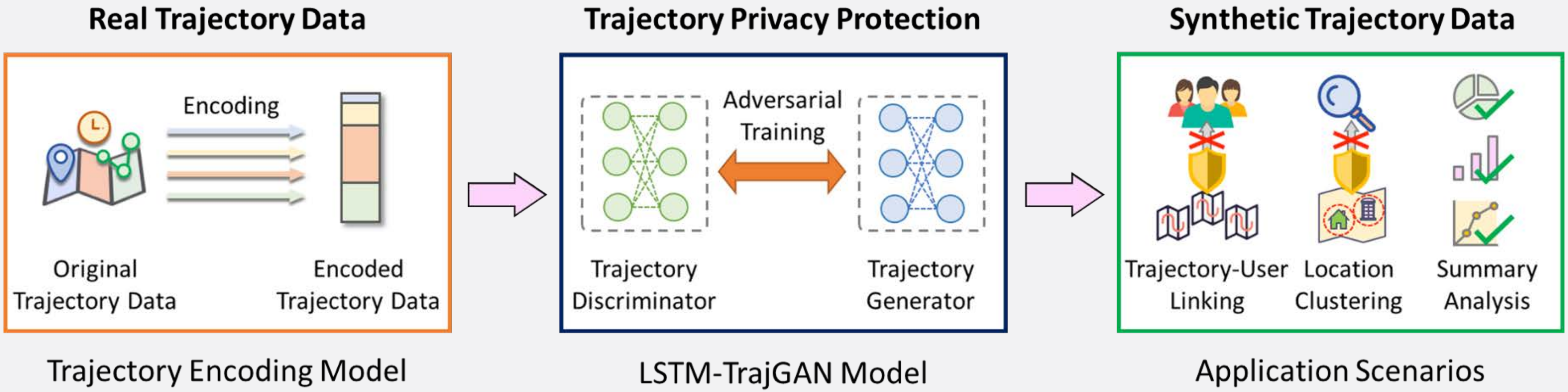


- Random Perturbation
- Donut Method
- Gaussian Method
- Voronoi Masking

Geomasking (IJGIS, 2023)

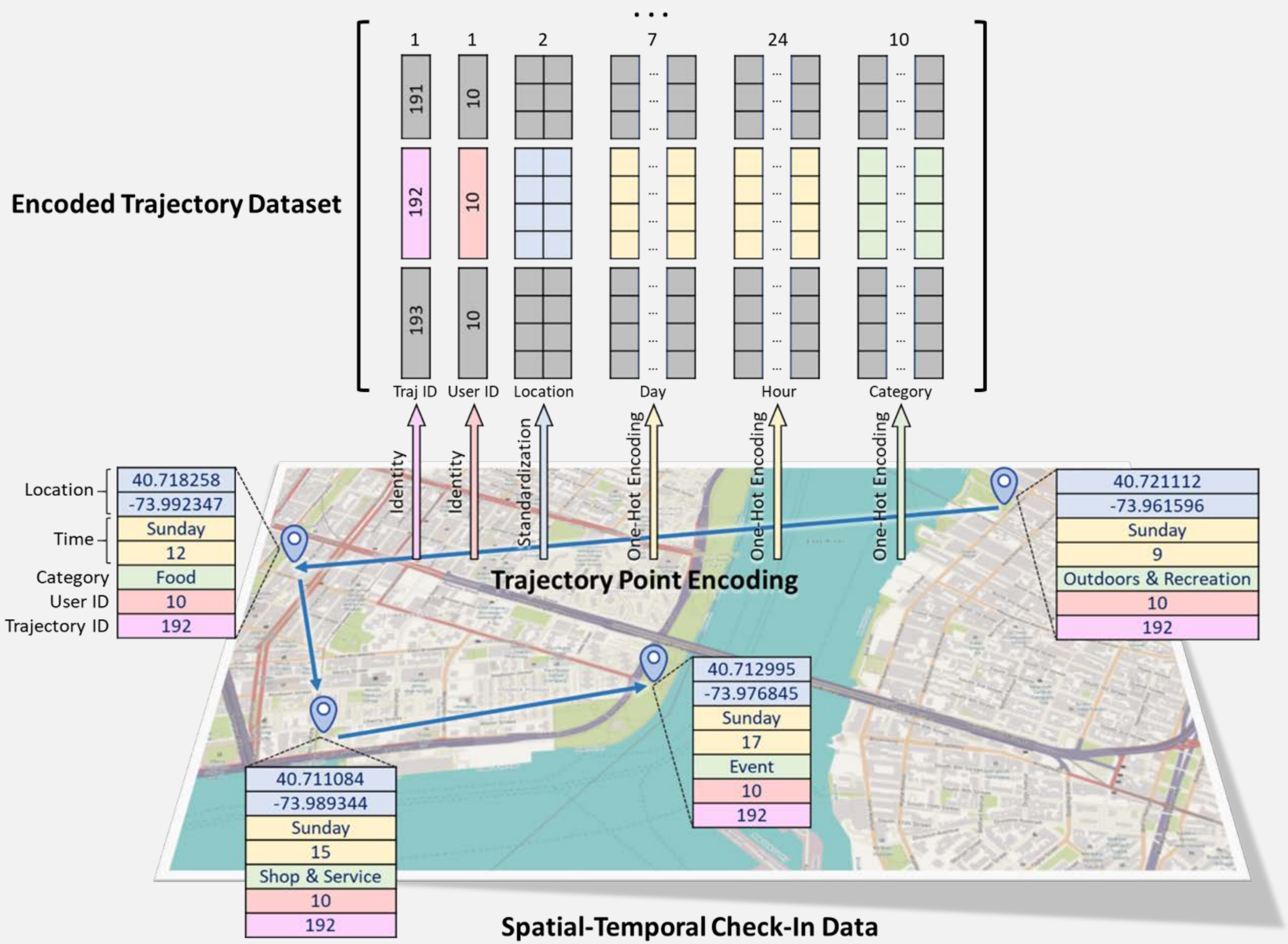
Overview of the LSTM-TrajGAN Model

- **RQ1: How effective is a generative GeoAI model using deep learning in protecting the trajectory creators from being re-identified? (i.e., privacy protection effectiveness)**
- **RQ2: Can the synthetic trajectories preserve the semantic features (spatial-temporal-thematic characteristics) compared to real trajectories? (i.e., utility)**

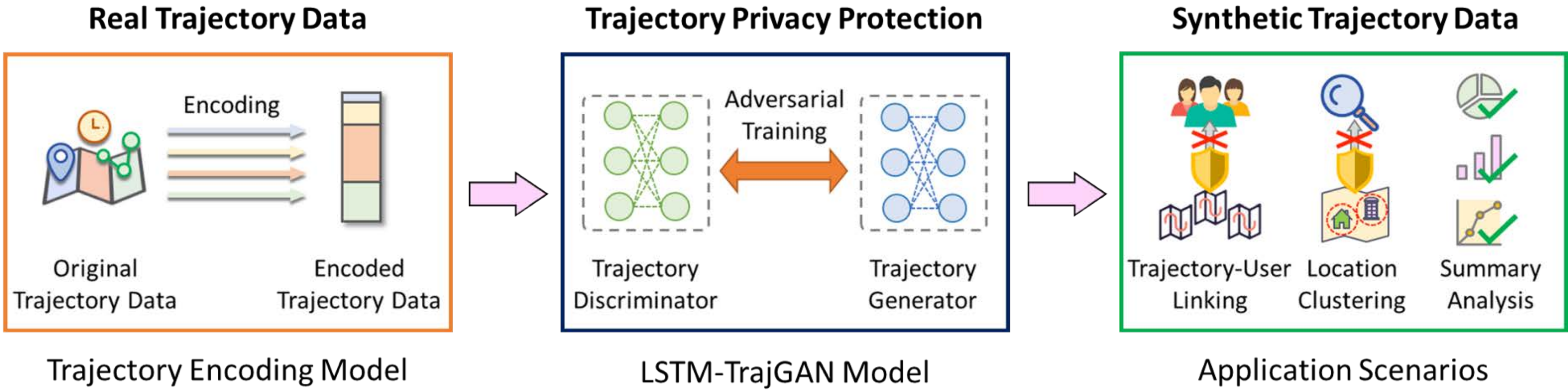


Rao, J., Gao, S., Kang, Y., & Huang, Q. (2020). LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. *In the Proceedings of the 11th International Conference on Geographic Information Science (GIScience 2021), No. 12; pp. 12:1–12:17.* DOI: 10.4230/LIPIcs.GIScience.2021.12

Trajectory Encoding Model



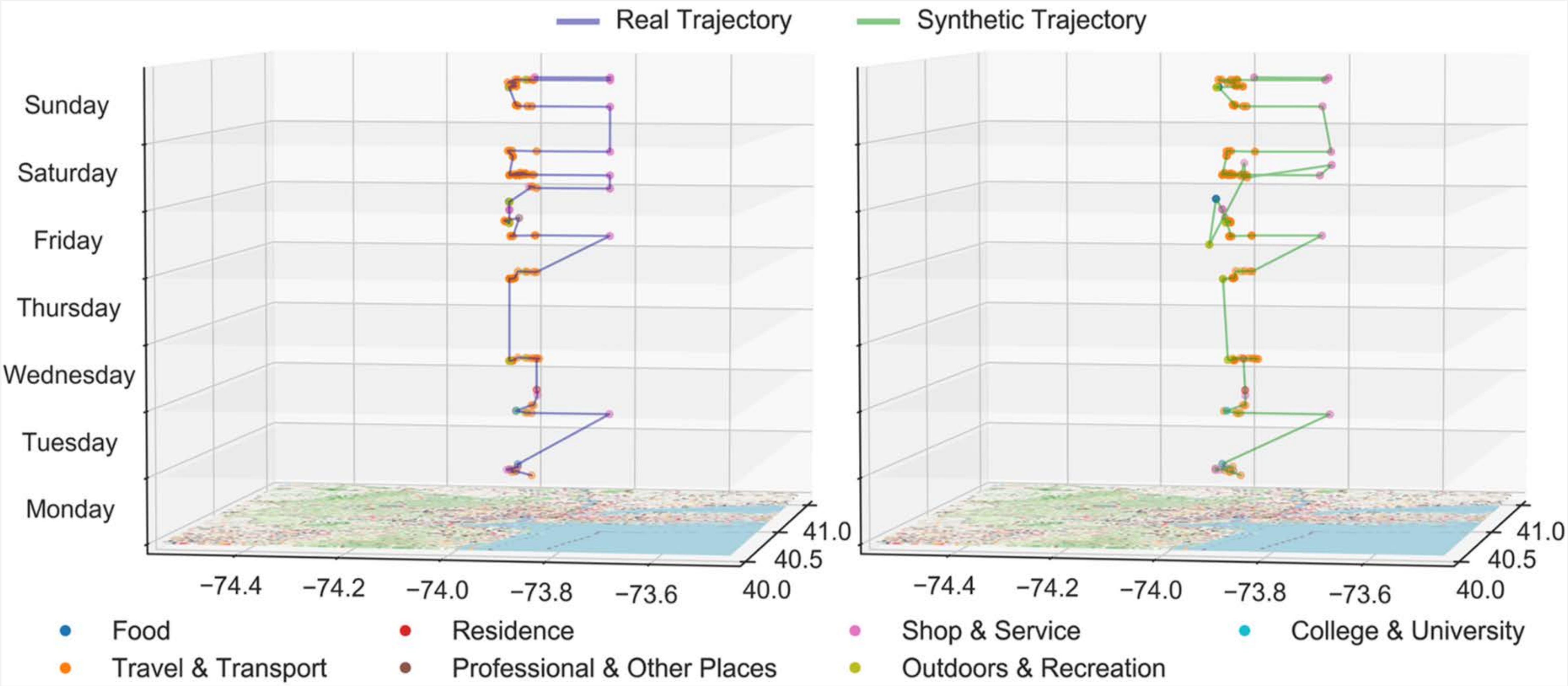
LSTM-TrajGAN Loss Function



$$TrajLoss(y^r, y^p, t^r, t^s) = \alpha L_{BCE}(y^r, y^p) + \beta L_s(t^r, t^s) + \gamma L_t(t^r, t^s) + cL_c(t^r, t^s)$$

https://github.com/GeoDS/LSTM-TrajGAN

Trajectory Generation



Trajectory Generation Results

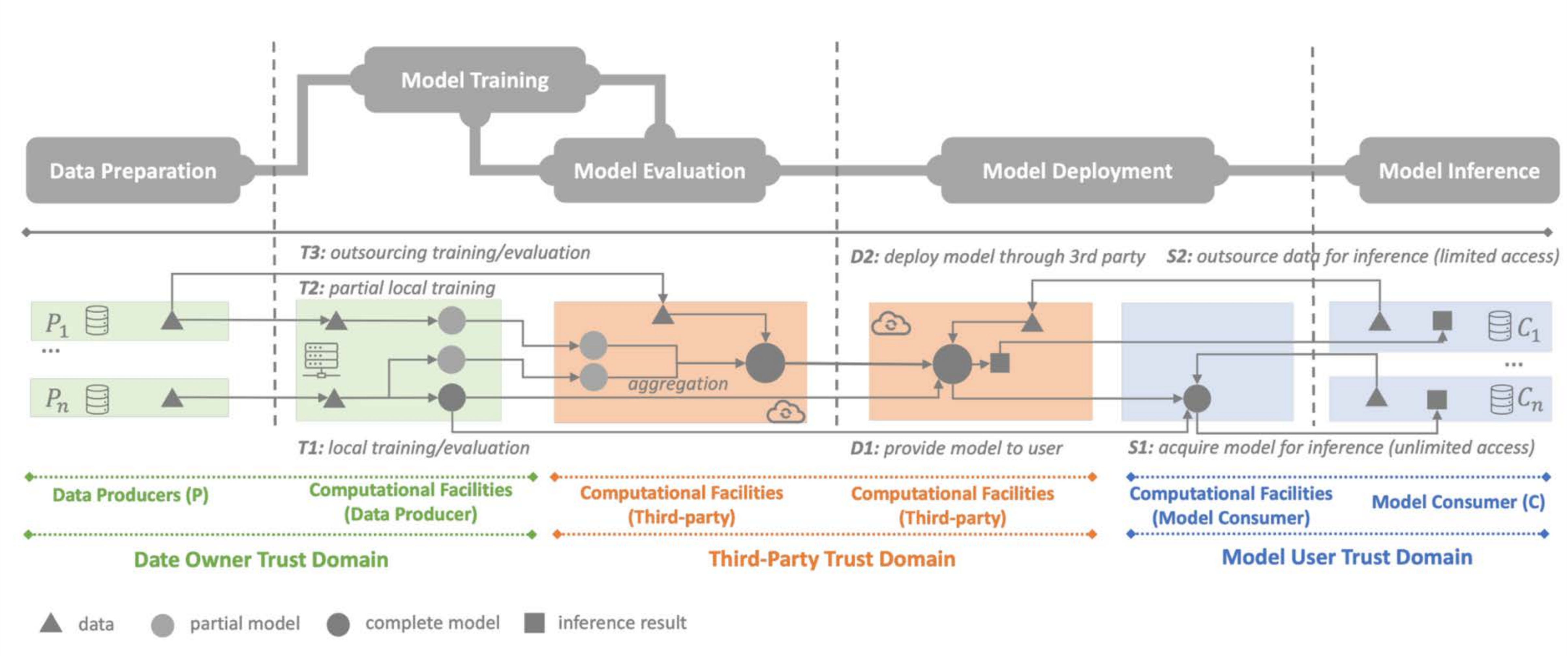
The privacy protection effectiveness of different privacy protection methods on the Trajectory User Linking task.

Method	ACC@1	ACC@5	Macro-F1	Macro-P	Macro-R
Original	0.938	0.976	0.925	0.937	0.927
RP (Spatial Only)	0.777	0.934	0.758	0.806	0.764
RP (Spatial-Temporal)	0.668	0.888	0.640	0.711	0.654
Gaussian (Spatial Only)	0.561	0.832	0.522	0.573	0.537
Gaussian (Spatial-Temporal)	0.486	0.766	0.431	0.488	0.470
LSTM-TrajGAN	**0.459**	**0.722**	**0.381**	**0.429**	**0.428**

Note: RP stands for Random Perturbation; Gaussian stands for Gaussian Geomasking

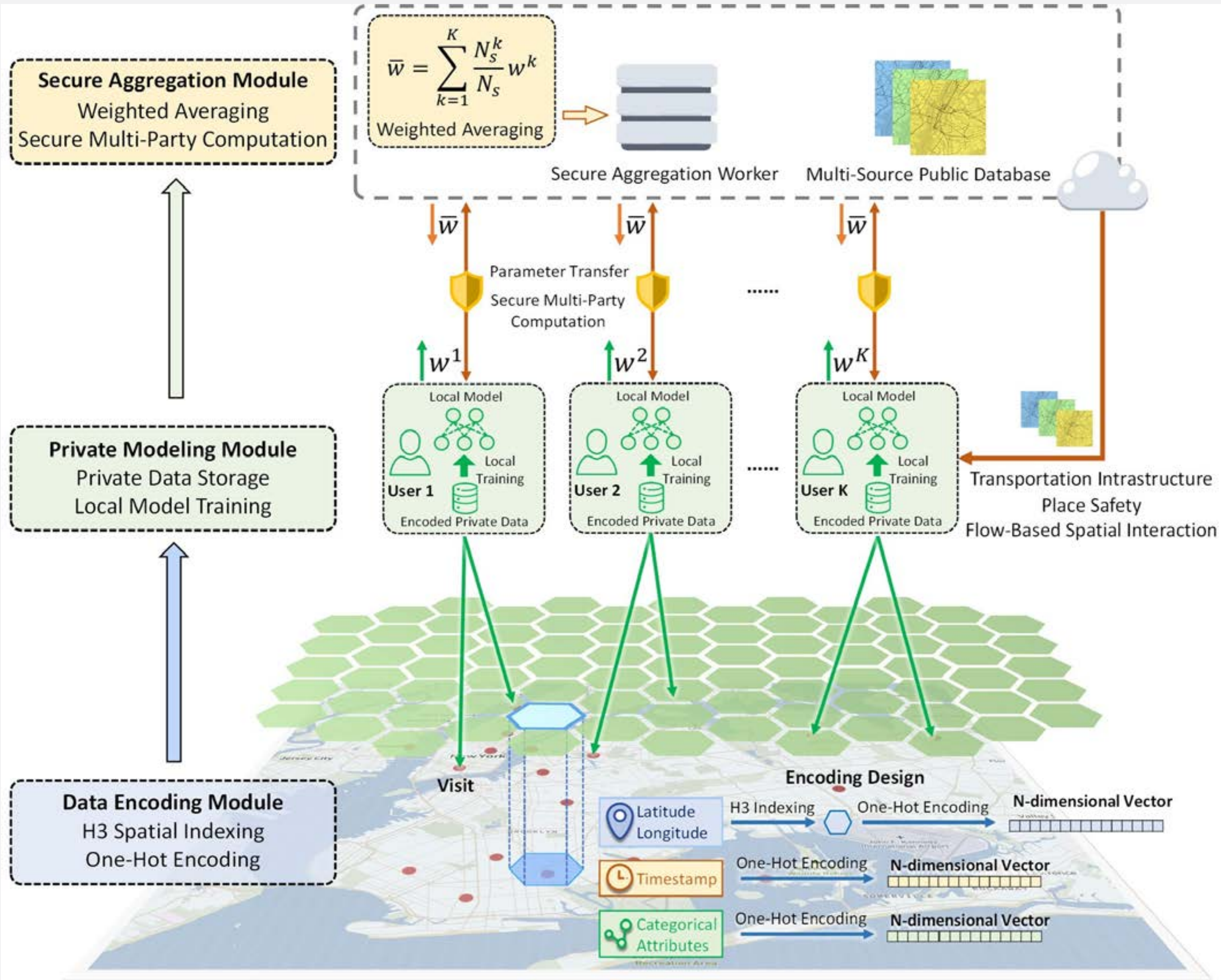
Privacy-Preserving Machine Learning

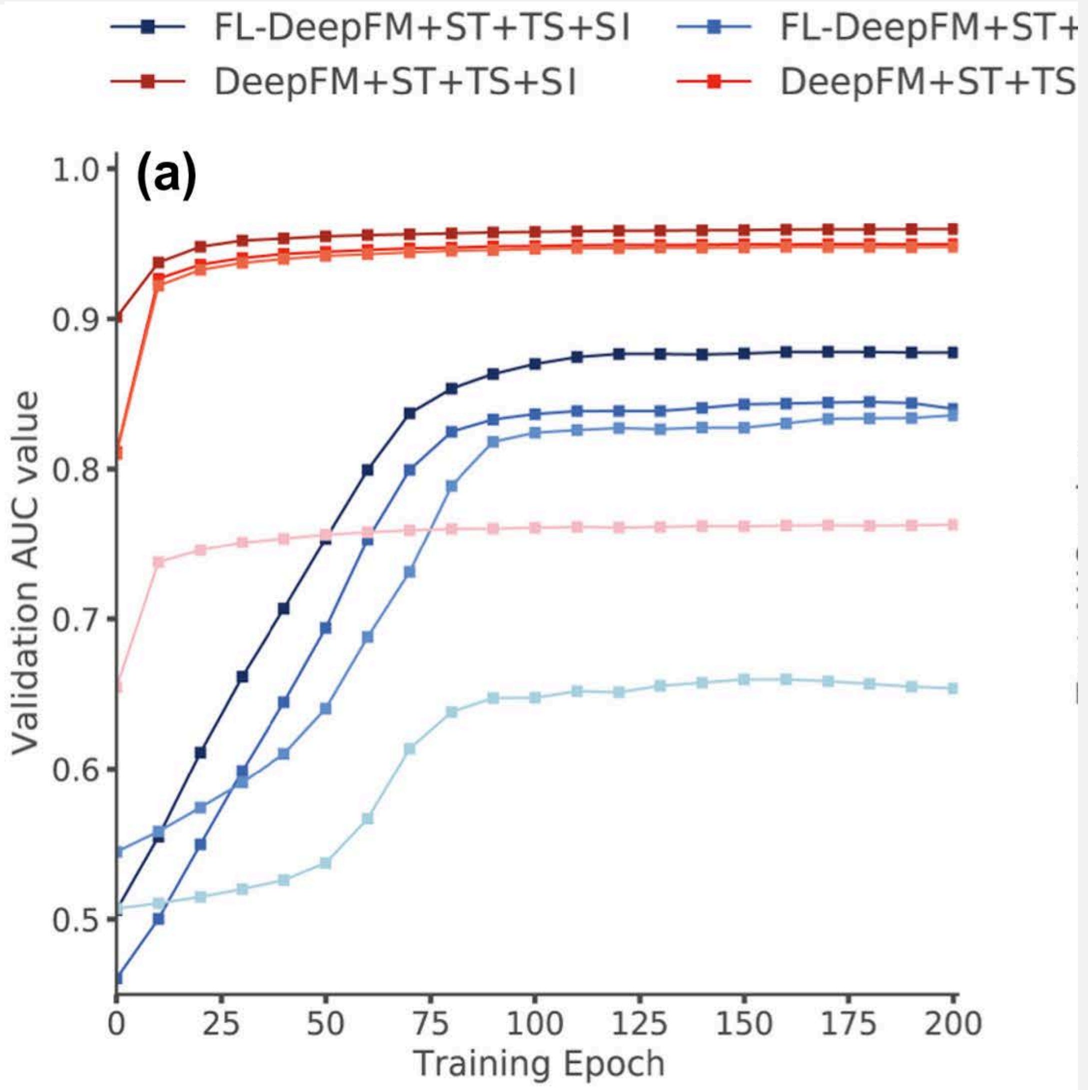
Build trust in model training and inference



(Xu, Baracaldo, and Joshi, 2021, arXiv)

Federated Learning-Based Framework





Rao, J., Gao, S., Li, M., & Huang, Q. (2021) A privacy-preserving framework for location recommendation using decentralized collaborative machine learning. *Transactions in GIS,* 25(3), 1153-1175.

Conclusion

- Understanding the regularity in human mobility patterns
- Protecting location privacy with GeoAI approaches

Thank you!

Geospatial Data Science Lab
UW-Madison

https://geods.geography.wisc.edu

Twitter: @gissong